Appl. No. 10/027,309
Amdt. dated March 11, 2005
Reply to final Office action of January 11, 2005

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1.-20. (Canceled).

21.   (Currently amended) A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing estimates for row. counts and unique entry counts for a singe column operator;

determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a second selectivity estimate based on an assumption that the columns are substantially dependent on each other;

determining a third selectivity estimate for predicates in the query using the first and second selectivity estimates, the third selectivity estimate being used in optimizing processing of the query by the database management system;.

determining a cross product from the single column statistics;

determining a measure of dependency; and

calculating the third selectivity estimate as the product of a difference between the first and second selectivity estimates plus one of the first or the second selectivity estimates.

22.   (Canceled).

23.   (Original) The method of claim 21, wherein the plurality of columns are substantially independent on each other.

24.   (Original) The method of claim 23, wherein the third selectivity estimate is substantially equal to the first selectivity estimate.

25.   (Original) The method of claim 21, wherein the plurality of columns are dependent on each other.

26.   (Original) The method of claim 25, wherein the third selectivity estimate is substantially equal to the second selectivity estimate.

27.   (Original) The method of claim 21, wherein the third selectivity estimate is within a range between the first and second selectivity estimates.

28.   (Original) The method of claim 27, further comprising determining an estimate of a dependency of the columns.

29.   (Original) The method of claim 28, wherein the estimate of the dependency of the columns is used to determine the third selectivity estimate.

30.   (Original) The method of claim 21, wherein the third selectivity estimate is chosen to be in a central range between the first and second selectivity estimates.

31.   (Canceled).

32.   (Previously presented) A method for optimizing a database management system process of a query, the method comprising:
    collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing estimates for row counts and unique entry counts for a singe column operator;
    determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a first factor as a measure of a skew of the plurality of columns and as a measure of a dependence of a plurality of the columns;

determining a second selectivity estimate for predicates in the query using the first selectivity estimate and the first factor, the second selectivity estimate being used in optimizing processing of the query by the database management system,

wherein the first factor is determined by

computing a product of unique entry count selectivities from a sum of maximum unique entry counts for the plurality of columns,

computing a product of maximum initial unique entry counts for the plurality of columns,

computing a ratio of the product of unique entry count selectivities and the product of maximum initial entry counts,

selecting a maximum multicolumn unique entry count from multicolumn entry counts for the plurality of columns, and

computing the first factor from a product of the ratio and an inverse of the maximum multicolumn unique entry count.

33.   (Currently amended) The method of claim 32~~31~~, wherein the plurality of columns are substantially independent on each other.

34.   (Original) The method of claim 33, wherein the second selectivity estimate is substantially equal to the first selectivity estimate.

35.   (Currently amended) The method of claim 32~~31~~, wherein the plurality of columns are dependent on each other.

Appl. No. 10/027,309
Amdt. dated March 11, 2005
Reply to final Office action of January 11, 2005

36.   (Currently amended) The method of claim 32~~31~~, wherein the second selectivity estimate is a product of the first factor and the first selectivity estimate.

37.   (Currently amended) A data processing system, comprising:
        a processor;
        a memory coupled to the processor;
        wherein the memory stores a compiler that, when executed by the
                processor, determines a join selectivity value of two columns
                based on a first selectivity value that assumes the two columns
                are dependent and a second selectivity value that assumes the
                two columns are independent, and
        wherein the compiler performs a join operation based on the join
                selectivity value,
        wherein the compiler determines an intermediate selectivity value
                approximately halfway between the first selectivity value and the
                second selectivity value when a dependence between the two
                columns is unknown and wherein the compiler performs a join
                operation based on the intermediate selectivity value.

38.   (Previously presented) The data processing system of claim 37 wherein the compiler determines the join selectivity of two columns further based on a cross product of row counts estimated for each of the two columns.

39.   (Previously presented) The data processing system of claim 38 wherein the row counts are estimated by a quantity of unique entry counts for each of the two columns.

40.   (Canceled).

BEST AVAILABLE COPY

Appl. No. 10/027,309
Amdt. dated March 11, 2005
Reply to final Office action of January 11, 2005

41.   (Currently amended) A storage medium containing computer-readable instructions that are executable by a computer and cause the computer to:

   produce a query tree based on a query posed by a computer language statement;

   transform the query tree into a form that represents a number of logically equivalent methods of processing the computer language statement;

   estimate a cost associated with carrying out each of the logically equivalent methods,

   wherein said estimate a cost comprises determining a join selectivity for two columns based on a first selectivity value that assumes the two columns are dependent, a second selectivity value that assumes the two columns are independent and, when an interdependence of the two columns is unknown, an intermediate value between the first selectivity value and the second selectivity value ~~and a cross product of row counts for each of the two columns~~.

42.   (Canceled).

43.   (Previously presented) The storage medium of claim 41 wherein said determining a join selectivity for two columns is further based on a skew calculation that provides a correction if the two columns have different row count to unique entry count ratios.